Proceedings of the 1991 Winter Simulation Conference
Barry L. Nelson, W. David Kelton, Gordon M. Clark (eds.)

RESEARCH ISSUES IN METAMODELING

Robert G. Sargent

Simulation Research Group
449 Link Hall
Syracuse University
Syracuse, NY 13244

ABSTRACT

This paper identifies and discusses some of the research issues in metamodeling of simulation models after giving an overview of the metamodeling process. Metamodeling of the M/M/1 queue is used to illustrate some of the issues.

1 INTRODUCTION

With enormous computing power at extremely low cost soon to occur (e.g. with parallel computers), the capability to generate large amounts of output data from simulation models cheaply and quickly will also occur This necessitates ways to effectively use this capability. One possible way is to use metamodels (Kleijnen 1987).

A metamodel here is a polynomial (mathematical) model that relates the input-output behavior of the simulation model as a "black box". (See Figure 1.) A metamodel is often a least squares regression model of the following form:

$$E[y] = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \ldots + \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_{ij} x_i x_j + \ldots \qquad (1)$$

where the β's are parameters, x's are input variables, and y is the output or response. The metamodel hopefully only contains the input variables that are necessary to describe the simulation model's response or output behavior over the experimental region. The objective of a metamodel is to "effectively" relate the output data of a simulation model to the model's input to aid in the purpose for which the simulation model was developed. We note that a simulation model is a causal (or mechanistic) model and a metamodel is an empirical model (Box and Draper 1987).

The purpose of this paper is to identify research issues in metamodeling. Hopefully, this will accelerate research on them. We note that simulation models are run on the computer which means we have complete

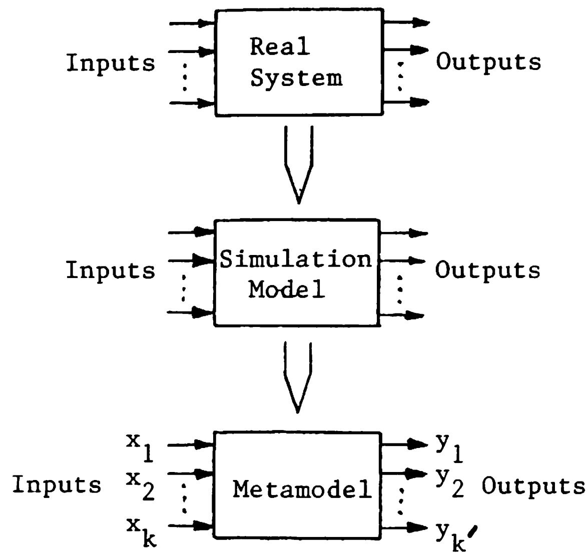

Figure 1: Modeling Relationships

control of a model and that these models usually have a large number of input (and internal) variables that can be controlled. Neither of these are usually true in other forms of experimentation. Also, we note that the issue of validity occurs since we are using models instead of working with the real system.

The remainder of this paper is organized as follows. In Section 2 we give an overview of the metamodeling process. Section 3 contains an example using the M/M/l queue to illustrate some of the issues in metamodeling. Section 4 identifies research issues. The summary is in Section 5.

2 OVERVIEW OF THE METAMODELING PROCESS

In this section we briefly present an overview of the metamodeling process (i.e. the methodology or steps involved) as we currently view it. We first give the properties required of specific metamodels that one may desire to develop. Next, we give the issues that must be decided in developing specific metamodels. These issues can be divided into strategic and tactical issues.

2.1 Underlying Properties

In this subsection we give the properties underlying a specific metamodel that one desires to develop.

The Properties:

P1. What is the purpose of the metamodel: system understanding, prediction - what is the value of the system output given values of the system inputs, system optimization, or approximating the response surface?
P2. Do we need a single response or multiresponse metamodel?
P3. Are the responses deterministic or random?
P4. Are unimodal or multimodal responses expected?
P5. How many input (and internal) variables are to be considered? Which variables are qualitative and which are quantitative?
P6. What is the experimental region (or region of applicability)?
P7. What amount of accuracy is needed in the metamodel?

2.2 Decision Issues

In this subsection we give the major decisions that must be made in developing a metamodel.

The Decisions:

D1. What type of metamodel should be used: least squares, Taylor series expansion, etc.?
D2. What type of criterion should be used to determine Goodness of Fit of the metamodel: maximum absolute error, relative absolute error, etc.?
D3. Should separate screening experiments be conducted to identify the input variables to be included in the metamodel? If yes, should sequential or simultaneous procedures be used? Also, what specific procedure (or type of experiment) should be used? How should the simulation response (output) data be generated?
D4. What type of experimental design should be used over the experimental region: full factorial design, fractional factorial design, etc.? What techniques should be used to generate the response data: independent runs, common random numbers, control variates, correlation-induction, etc.? If the simulation model is random, how should the (usual) requirements of independence, normality, and constant variance be satisfied?
D5. Does the metamodel have the necessary accuracy required? If not, what actions are required?
D6. How is the validity of the metamodel with respect to the simulation model determined?
D7. How is the validity of the metamodel with respect to the system determined?
D8. Is sensitivity analysis required? If so, how?

The decision issues can be divided into strategic and tactical issues. The strategic issues have to do with determining the answer to the basic decisions in D1-D8 and also whether they should be integrated, e.g. should D3 and D4 be combined into one step of the metamoldeling process. Tactical issues have to do with "the specifics" once the strategical decisions have been made, e.g., how long should run lengths be and how should transient data be eliminated in steady-state simulations.

3 EXAMPLE

We will use the M/M/1 queue with FIFO queue discipline and infinite queue capacity to illustrate some of the research issues in metamodeling. The performance measure (output variable) of interest is the expected steady-state sojourn time, W, in the system (i.e. waiting plus service time). Let $\lambda = 0.1$ be the arrival rate and $S = 1/\mu$ the mean service time. We are interested in finding a metamodel to describe the relationship between W (the expected response) and S over the range $5 \leq S \leq 9$. Values of the true response W for different values of S are well known and are given by $W = S/(1 - \lambda S)$. (See Figure 2).

The first case we will consider is to develop a least squares metamodel using values from the known relationship between W and points chosen in the experimental region of $5 \leq S \leq 9$. Using the exact values is like having a deterministic simulation. (We note that with enormous computing power at extremely low cost that it will often be possible to obtain point estimations of mean responses of stochastic simulations such that they will have almost no variance--this implies that it may be possible to fit metamodels to these responses like metamodels to deterministic simulations.) The first through the fifth order metamodels are contained in Table 1 and the first and second order metamodels are shown in Figure 2. Table 1 also contains the "Goodness of Fit" criterion Maximum Absolute Error (MAE), maximum $| W(S) - W'(S) |$ over the experimental region, where W(S) and W'(S) respectively denote the exact and metamodel predicted values of the expected steady-state sojourn times for a given value of S. One readily sees from Table 1 and Figure 2 that a metamodel higher than a second order is needed if a "good approximation" is desired. (We note that books and articles are devoted mainly to developing first and second order metamodels.)

The second case we consider is to use a Taylor series expansion of the true relationship as a metamodel,

$$W = \sum_{j=0}^{q} \beta_j (S - S_o)^j \qquad (1)$$

where $\beta_o = S_o/(1 - \lambda S_o)$, $\beta_j = \lambda^{j-1}/(1 - \lambda S_o)^{j+1}$, $j = 1, 2, \ldots, q$ and S_o is the value of S at which the expansion occurs. If we expand at the midpoint of the experimental region, $S_o = 7$, the β_j's values are given in Table 2 along with the MAE's and Figure 2 contains the first order model. We observe differences between the least squares and Taylor series metamodels. We first observe that the least squares metamodels have less MAE for a specific order. Secondly, we observe that at the point of expansion, the Taylor series metamodels has zero error;

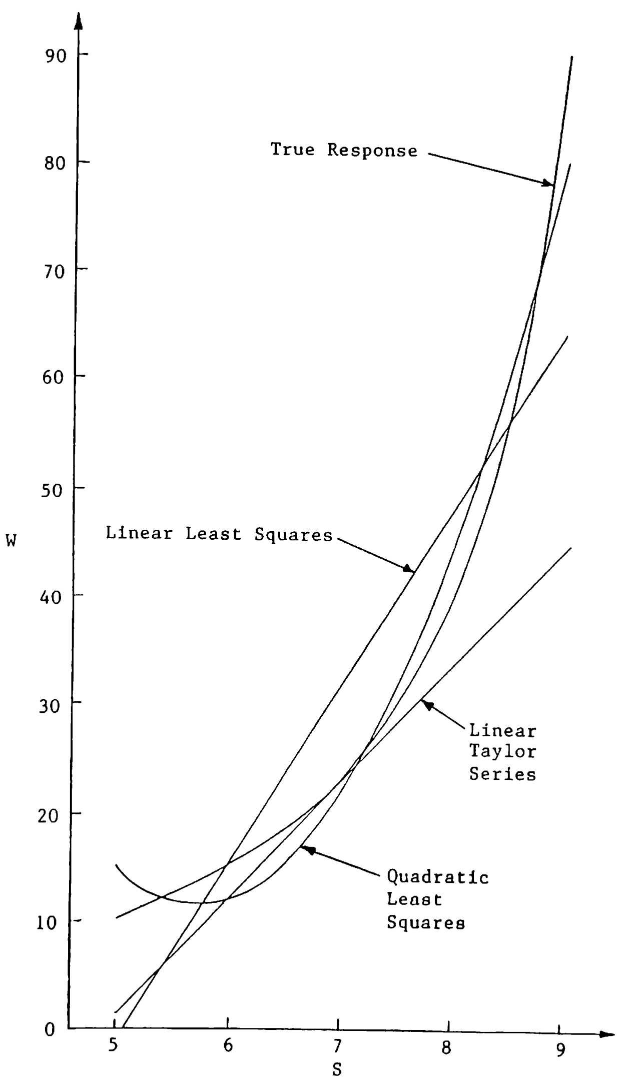

Figure 2: Different Metamodels

Table 1: Different Least Squares Metamodels for Steady-State M/M/1 Sojourn Times Using Data From the Exact Analytical Relationship

Metamodel	Order	MAE
$W = -83.04 + 16.36S$	1st	25.7
$W = 220.05 - 72.96S + 6.38S^2$	2nd	9.8
$W = -578.92 + 282.72S - 454.85S^2 + 2.47S^3$	3rd	3.4
$W = 1527.84 - 970.69S + 230.40S^2 - 2.414S^3 + 0.95S^4$	4th	1.1
$W = -3930.55 + 3094.34S - 967.44S^2 + 150.50S^3 - 11.65S^4 + 0.36S^5$	5th	0.5

whereas, in general, the lower order least square metamodels have errors. Thus, which type of metamodel to use may depend upon the purpose of the metamodel. We observe two other properties of the Taylor series metamodels. One is that expanding at the midpoint of the experimental region often will not give the "best" fit over the total experimental region; e.g. expanding about $S_o = 7.4$ gives smaller MAE's than expanding about $S_o = 7.0$. Another property of the Taylor series metamodels is that they may not converge to the true value over the experimental region; e.g. if we expand about $S_o = 8.0$, the Taylor series metamodel will not converge to the true value at, e.g., $S_o = 5.0$. Thus, research is needed on when to use what type of metamodel. A related issue is what type of criterion should be used for Goodness of Fit.

The third case we will consider is to fit a least squares regression model to data generated by simulation. We will use a straight forward approach (i.e. we will not use, e.g., variance reduction techniques). The required assumptions on the data were satisfied as follows:

Table 2: Taylor Series Expansion Metamodels for Steady-State M/M/1 Sojourn Times

j	β_j	MAE*
0	23.333	66.7
1	11.111	36.6
2	3.704	29.6
3	1.235	19.8
4	0.412	16.5
5	0.137	8.8
6	0.046	5.9
7	0.015	3.9

*For order j

Table 3: Experimental Conditions and the Derived Metamodels for Steady-State M/M/l Sojourn Time Using Simulation and Regression Analysis

S	Truncation Point	Run Length	Effective Run Length
5	25	75	50
6	50	190	140
7	120	640	520
8	300	3,300	3,000
9	1,500	55,500	54,000

Metamodel	Order	R^2
$W = -77.46 + 16.36S$	1st	0.760
$W = 250.11 - 81.22S + 6.97S^2$	2nd	0.958
$W = 138.53 - 30.98S - 0.38S^2 + 0.35S^3$	3rd	0.959

(i) independence of responses generated on different runs was ensured by using independently seeded runs; (ii) constant response variance across runs was ensured by choosing appropriate run lengths using the guidelines in Whitt (1989); (iii) initialization bias was eliminated by selecting truncation points slightly larger than given by the guidelines in Whitt (1991); and (iv) normality of the final response generated on each run was obtained by averaging the observations over each run to obtain one data point per run. The experimental conditions and the first, second, and third order metamodels are contained in Table 3 along the Goodness of Fit criterion R^2 which denotes the squared multiple correlation coefficient. We note that R^2 must approach 1 as the number of estimated parameters in a metamodel approach the number of experimental points (runs) and that this is not a satisfactory goodness of fit criterion. Some research issues are how to satisfy the assumptions required on the data. In general, we do not know how to satisfy assumptions (ii) and (iii) and these are major research issues. (See, e.g., Welch 1990.)

In the approaches used in cases one and three, we must select experimental points to obtain our data values. We typically use classical experimental designs for this. However, in computer based experiments one has complete control of a model and thus has complete freedom in performing experiments. Most experimental designs are for physical experiments which have constraints. Research is needed for experimental designs for both deterministic (case one) and stochastic (case three) simulations.

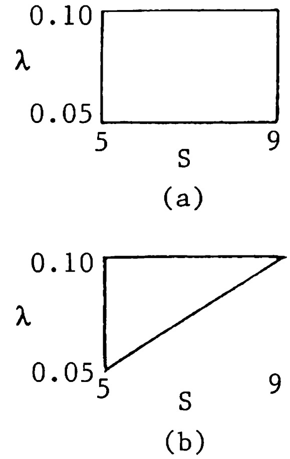

Figure 3: Experimental Regions

Furthermore, in case three we also have statistical considerations to be concerned with and can be used for efficiency, e.g. variance reduction techniques can be used. Examples of some research here is Donahue, Houck, and Myers (1990) and Schuben and Margolin (1978). We note that the limited research to date has been devoted primarily to first and second order metamodels. A related topic is how are the statistical assumptions required of the metamodel satisfied. Almost no research work has been performed on this; an exception is Tew and Wilson (1990).

Let us next consider two other dimensions of metamodeling of the M/M/l queue. First, suppose we wanted a metamodel where both λ and S were input variables instead of only S. Almost all experimental designs are based on rectangular experimental regions as in Figure 3(a). However, often in practice this is not the case; e.g., one might be interested in a metamodel over the experimental region given in Figure 3(b).

Another dimension of metamodeling of the M/M/l queue is when one is interested in more than one performance measure (output variables); e.g. suppose the performance measures included the utilization of the server and the percentage of customers that had to wait over five minutes in addition to the expected steady-steady sojourn time. Almost no work exists on how to develop a multiresponse metamodel instead of developing three separate metamodels. One exception is Porta Nova and Wilson (1989).

4 RESEARCH ISSUES

This author believes that there are major research issues regarding each Decision Issue (which includes both strategic and tactical issues), and also how the Decision Issues interact with the Properties. Some of these should be clear from the example in Section 3. We identify some of the research issues here.

Some Research Issues:

- Rl. What type of metamodel to use with respect to the properties, in particular Pl, needs to be determined?
- R2. What type of Goodness of Fit criterion should be used with respect to the properties and type of metamodel used?
- R3. Under what conditions should screening experiments be conducted in order to determine the input variables to be used in developing a metamodel? It is known that sequential experimentation reduces the number of experiments (runs) that are needed as compared to simultaneous experiments (e.g. see Wilde 1964). How should this approach be used in metamodeling? Recently, Bettonvil and Kleijnen (1991) gave one such procedure. (Also, see Baily, Bauer, and Marsh 1989.) How to generate the response data efficiently (tactical issues) is another set of research issues (see R4).
- R4. The type of experimental design to use is a major research area. It is related to all of the properties, many of the decisions, some of the research issues, and the tactical issues in generating the data. Of specific interest is what are good designs for higher order metamodels. What are the trade-offs between simplicity and complexity (using, e.g., variance reduction techniques) in designs. Can sequential procedures be developed that includes screening. Tactical issues for stochastic simulation, in particular how to obtain constant variance, need research. Robustness to assumptions seem very important. Different shape experimental regions need to be considered.
- R5. Metamodels needs to have a certain amount of accuracy. If they do not, can efficient methods be developed to modify a metamodel to obtain the desired accuracy. An example of research is to identify which input variables or interaction terms need to be added.
- R6. A metamodel needs to be a valid representative of a simulation model. How should this be accomplished? Kleijnen (1987) provides some discussion but research is needed.
- R7. There is almost nothing in the metamodeling literature on determining the validity of the metamodel with respect to the real system. This is critical and is a research area that should receive high priority.
- R8. The relationships among the system, simulation model, and metamodel, including the model accuracies and validities, need research.
- R9. How does one perform sensitivity on a metamodel is a research issue.
- R10. Research is needed on how metamodels can be used to obtain greater understanding of systems.
- R11. Should transformations be used on input and output variables and if so, how, is a research topic.

5 SUMMARY

We attempted to identify some of the major research issues in metamodeling. This author believes that a "larger" view of metamodeling is needed in order to identify all of the research issues. It is hoped that this paper will stimulate discussions on and research in metamodeling. With the expectation of enormous inexpensive computing power in the very near future, effective ways need to be found to use this capability and metamodeling may be one such way.

ACKNOWLEDGMENTS

The author appreciates the discussions he has had on metamodels with Stephen Hsiao, Tapas Som, and Jim Wilson. Also, Tapas Som provided computer assistance for the example.

REFERENCES

Bailey, T.G., K. Bauer, and A.B. Marsh. 1989. Response Surface Analysis of Stochastic Network Performance, *Proceedings of the 1989 Winter Simulation Conference*, 437-443.

Bettonvil, B. and J.P.C. Kleijnen. 1991. Identifying the Important Factors in Simulation Models with Many Factors, Report FEW 498, Department of Economics, Tilburg University, Tilburg, the Netherlands.

Box, G.E.P. and N.R. Draper. 1987. *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, New York.

Donohue, J.M., E.C. Houck, and R.H. Meyers. 1990. Some Optimal Simulation Designs for Estimating Quadratic Response Surface Functions, *Proceedings of the 1992 Winter Simulation Conference*, 337-343.

Kleijnen, J.P.C. 1987. *Statistical Tools for Simulation Practitioners*, Marcel Dekker, New York.

Porta Nora, A.M. and J.R. Wilson. 1989. Estimating Multiresponse Simulation Metamodels Using Control Variates, *Management Science*, 35, 11,